August 12, 2024

Hon. Ona T. Wang
United States Magistrate Judge
United States District Court
Southern District of New York

Re: *The New York Times Company v. Microsoft Corporation, et al.,*
Case No.: 23-cv-11195-SHS-OTW: Dispute Re: OpenAI's Custodians

Dear Magistrate Judge Wang:

Plaintiff The New York Times Company ("The Times") requests a conference to discuss its request for OpenAI to add 17 custodians, each of whom likely has "non-duplicative documents." *U.S. Bank v. Triaxx Asset*, 2021 WL 3024824, at *1 (S.D.N.Y. July 16, 2021); *Fort Worth Employees v. JPM*, 297 F.R.D. 99, 106 (S.D.N.Y. 2013) (granting 18 more custodians).[1]

1.   OpenAI Refuses to Add Additional Custodians Proposed by The Times

Since June 5, The Times has resisted OpenAI's efforts to artificially limit custodians, first to 10 and later to 12. Dkt. 135. On June 14, The Times made a proposal for 24 additional custodians. OpenAI agreed to just 4—those identified in Rule 26 disclosures and Interrogatories.

OpenAI has failed to identify custodians for all of the issues in this case, which is particularly improper in light of its obligations in other cases. In the *Tremblay* case in N.D. Cal., a case its co-Defendant Microsoft has described as far narrower, OpenAI was ordered to provide an "initial" set of 24 custodians, with plaintiffs permitted to seek more. 23-cv-03223, Dkt. 166; *see* Dkt. 48 at 6 (Microsoft claiming this case "raises the issue of generative AI outputs to a degree not seen in any of the other cases, either in California or New York"). OpenAI's productions to date skew heavily toward model training, omitting many documents related to generative search, model outputs, and other relevant issues. Because OpenAI's initial custodian list has yielded productions slanted toward training, OpenAI may have recycled the same custodians it is using in the *Authors Guild* class action pending before this Court, a case OpenAI has described as "very different" because it is "narrowly focused" on training. Dkt. 72 at 16-17. Finally, any concerns about burden should be disregarded because, in violation of the agreed portion of the ESI Order, OpenAI has refused to provide search term hit counts for the proposed custodians. Ex. 1 at 10; Dkt. 135-2 ¶ 5.

2.   Proposed Custodians in Leadership Roles for Central Issues

The Times seeks documents from several high-level OpenAI employees in possession of documents central to the issues of this case. *See, e.g., Shenwick v. Twitter, Inc.*, 2018 WL 833085, at *1 (N.D. Cal. Feb. 7, 2018) (adding Twitter's CEO as a custodian).

Greg Brockman, OpenAI's former President and a Co-Founder, led OpenAI's efforts to

---

[1] The parties conferred by videoconference on July 15 and July 22, and exchanged many correspondences. Ex. 1.

Mira Murati is the Chief Technology Officer, and she has a unique role with respect to

██
██
██
██
██

Brad Lightcap is the Chief Operating Officer, with responsibility for █████████

██
██
██

Ilya Sutskever was a Co-Founder and the Chief Scientist. He has a unique perspective on Defendants' models and products, ███

██
██ Sutskever recently left OpenAI to form a new AI firm focused on "safe superintelligence."[3]

Sam Altman, a Co-Founder and the CEO, has driven key decision making relevant to this case, including █████

██
██
██
██
██
██

3. Proposed Custodians with Key Knowledge About Important Issues

None of the current custodians adequately address the below issues.

Retrieval Augmented Generation ("RAG"): OpenAI should add **Alex Passos,** ████████
██████████████████████████████████ Defendants argue that RAG makes this case "substantially broader" than *Authors Guild* (Dkt. 153 at 2-3 (Microsoft); *see also* Dkt. 72 at 16-17 (OpenAI claiming discovery will "differ substantially" from *Authors Guild*, including with respect to model "outputs")) – yet OpenAI has proposed no custodian for RAG, the process by which its GenAI products retrieve and copy publishers' content to generate responses to users. Passos ███
██

---

[2] *See* https://www.youtube.com/watch?v=s-KCt0Z47TQ at 1:02 (Lightcap discussing OpenAI's plans to allow publishers to "specify specifically how they'd like their content used with respect to AI training and further use").
[3] https://www.axios.com/2024/06/19/ilya-sutskever-openai-co-founder-safe-superintelligence.

█████████████████████████ OpenAI should add two employees who possess documents to support The Times's ██████████████████████████ **Alex Gray** has █████

██████████████████████████████████ **Ryan Lowe** is a former research manager who

██

██

██

██

██████ Failure to Address ████████████████ OpenAI should add the following three custodians █████████████████████████████████

█████████████████████████ . **Bianca Martin** is a researcher

█████████████████ **William Fedus** is a research team leader who

█████ **Johannes Heidecke** has information about a █████

██

██

Training Datasets: OpenAI should add three employees with knowledge about how the ██████████████ **Wojciech Zaremba** is a Co-Founder of OpenAI who has written about how a █████████████

██

██████████ **James Betker** is a research engineer who has written about how "model behavior is not determined by architecture, hyperparameters, or optimizer choices. It's determined by your dataset, nothing else."[4] **Chris Berner** is the Head of Compute, and according to LinkedIn, leads the team that "build[s] our supercomputing clusters and conduct[s] research in highly efficient model training."

OpenAI's Licensing Negotiations with The Times: OpenAI should also add **James Dyett**, the Head of Platform Sales █████████████████████████████████

███

OpenAI's Irresponsible Conduct Commercializing GenAI Products: Finally, OpenAI should add <u>two</u> former employees who have publicly criticized OpenAI's irresponsible behavior in the GenAI space, including with respect to the commercialization of Defendants' products, an issue relevant to fair use: **Daniel Kokotajlo** and **William Saunders**.[5]

Respectfully,

/s/ Ian B. Crosby

---

[4] https://nonint.com/2023/06/10/the-it-in-ai-models-is-the-dataset/.

[5] https://www.lesswrong.com/users/daniel-kokotajlo (Kokotajlo stating he "los[t] confidence that [OpenAI] would behave responsibly around the time of AGI"); https://www.businessinsider.com/former-openai-employee-williams-saunders-artificial-intelligence-building-titanic-apollo-2024-7 (Saunders stating that "when [OpenAI] release[s] things, their priorities are more like a product company. And I think that is what is most unsettling").

Ian B. Crosby
Susman Godfrey L.L.P.

/s/ Steven Lieberman
Steven Lieberman
Rothwell, Figg, Ernst & Manbeck

cc: All Counsel of Record (via ECF)